A Unified Framework for Fuzzy Spatio-Temporal Representation and Reasoning

Emmanuel Stefanakis
Department of Geography, Harokopio University,
70 E.Venizelou Ave., 17671 Athens, Greece.
e-mail: estef@hua.gr

Abstract. Recently, the interest has been focused on dynamic applications with geographic reference. These applications are commonly called as spatio-temporal applications and examine phenomena, which occur in specific geographical regions and change over time. In addition, data collected to model and analyze a phenomenon are accompanied with uncertainty. Current GIS technology turns out to be inefficient to model and handle the complex geographic phenomena, which involve space, time and uncertainty dimensions. In an effort to generate more intelligent GIS tools, this paper concentrates on representation and reasoning issues of geographic entities and introduces a framework with enhanced functionality.

Keywords. Space-time, uncertainty, fuzzy set methodologies, geographic entities, modeling, reasoning.

1 Introduction

Multicriteria decision analysis and a wide range of related methodologies, offer a rich collection of techniques and procedures to support decision making in traditional Information Systems. Recently, the interest has been focused on dynamic applications with geographic reference. These applications are commonly called as spatio-temporal applications and examine phenomena, which occur in specific geographical regions and change over time.

The major fruit of technological evolution on this area are geographic information systems (GIS), which although rich toolboxes, they have limited capabilities in modeling and handling complex spatio-temporal phenomena. Specifically, it is widely recognized that these systems (Leung and Leung 1993, Fischer 1994, Stefanakis 1997): do not accommodate the temporal dimension of geographic data; they provide a limited built-in analytical and modeling functionality; and their level of intelligence is inadequate. In addition, data collected to model and analyze a phenomenon are accompanied with uncertainty. The uncertainty is an inherent feature of geographic data and may arise through (Leung and Leung 1993): a) incomplete information associated to them; b) the presence of varying concentrations of attributes; and c) the use of qualitative descriptions of their attribute values and relationships. Current GIS applications usually ignore the uncertainty of data due to the inappropriate logical foundation (Boolean logic) of the GIS package used. Fuzzy set theory apparently provides a promising logical foundation for an *Intelligent GIS* (Stefanakis 1997, Stefanakis et al. 1999).

The scope of this paper is to introduce a framework with enhanced capabilities, in both representation and reasoning of geographic data. The discussion is organized as follows. Section 2 presents the dimensions of geographic entities. Section 3 presents briefly the fuzzy set theory concepts. Section 4 introduces a model to represent those entities, which is capable to accommodate all their dimensions. Section 5 describes the steps that should be followed in order to generate a database based on the proposed data model. Section 6 describes the properties of the operation model and focuses on the data interpretation operations applied on geographic entities. Representative operations (such as selection and overlay) are discussed in more detail. Finally, Section 7 concludes the discussion by summarizing the contribution of the paper and giving directions for future research.

2 Dimensions of Geographic Data

Geographic data have a set of characteristics that make them distinctly different from the more familiar lists and tables of alphanumeric data used in traditional business applications. Geographic entities (i.e., geographic objects, units or phenomena) have six dimensions along which attributes may be measured (Stefanakis 1997). These dimensions are (Figure 1):

1. *Identifier*: It provides a means to refer to different entities. Geographic entities may be assigned arbitrary names, geographic/place names, temporal or numeric codes. It is important for the naming system to be precise, consistent and able to guarantee uniqueness.

2. *Spatial data*: They describe the spatial characteristics of geographic entities. Spatial characteristics describe: a) the geographic position of an entity; b) its geometric characteristics; c) its presentational/mapping properties; and d) its spatial relationships (topological, direction, etc.) to other entities.
3. *Thematic data*: They describe the non-spatial (textual/numerical) properties of geographic entities (e.g., vegetation type, land parcel owner's name, highway class, etc.). This dimension heavily depends on the application domain.
4. *Temporal data*: They describe the temporal characteristics of geographic entities. Temporal characteristics describe: a) the temporal position of an entity; b) its temporal behavior, c) its temporal representation/mapping properties, and d) its temporal relationships (temporal topology, etc.) to other entities.
5. *Data quality*: It is a measure of the quality of data collected to describe geographic entities. It includes the spatial, temporal and thematic accuracy attributes assigned to geographic entities, the completeness and logical consistency among data elements.
6. *Multimedia data*: They consist of sound, pictures, vrml scripts and/or video data that accompany geographic entities (e.g., Greece should be accompanied by the national anthem, several pictures and videos of archaeological sites and islands).

Fig. 1 The six dimensions of geographic entities.

3. Fuzzy Set Theory Concepts

Fuzzy set theory (Zadeh 1965) is an extension of the classical set theory. A fuzzy set A is defined mathematically as follows: If $Z = \{z\}$ denotes a space of objects, then the fuzzy set A in Z is the set of ordered pairs:$A = \{z, MF^F_A(z)\}, z \in Z$, where the membership function $MF^F_A(z)$ is known as the "*degree of membership (d.o.m.) of z in A*". Usually, $MF^F_A(z)$ is a real number in the range [0,1], where 0 indicates no-membership and 1 indicates full membership. Hence, $MF^F_A(z)$ of z in A specifies the extent to which z can be regarded as belonging to set A. The choice of the membership function, i.e., its shape and form, is crucial and strongly affects the results derived by the decision-making process.

Fuzzy set theoretic *operations* provide the counterpart operations to those of classical set theory. In other words, logical operations with fuzzy sets are generalizations of usual Boolean algebra applied to observations that have partial membership of more than one set. The standard operations of *union*, *intersection*, and *complement* of fuzzy sets A and B, defined over some domain C, create a new fuzzy set whose membership function is defined as:

Union: $$MF^F_{A\cup B}(z) = \max\{ MF^F_A(z), MF^F_B(z) \}, \forall z \in C \quad (1)$$
Intersection: $$MF^F_{A\cap B}(z) = \min\{ MF^F_A(z), MF^F_B(z) \}, \forall z \in C \quad (2)$$
Complement: $$MF^F_{A'}(z) = 1 - MF^F_A(z), \forall z \in C \quad (3)$$

Consider the classification of individual locations on a layer based on the slope values with lexical values [*level, gentle, moderate, steep*]; and a second classification based on the land moisture with lexical values: [*dry, moderate, wet, water*].

For each individual location l (e.g., d.o.m. for level = 0.8 and d.o.m. for dry = 0.4) the d.o.m. value which provides an overall measure regarding: a) level ground *and* dry land is derived by: $\min\{MF^F_{level}(l), MF^F_{dry}(l)\}$, (e.g., $\min\{0.8,0.4\}=0.4$); b) level ground *or* dry land is derived by: $\max\{MF^F_{level}(l), MF^F_{dry}(l)\}$, (e.g., $\max\{0.8,0.4\}= 0.8$); and c) *non*-level ground is derived by: $1-MF^F_{level}(l)$, (e.g., 1-0.8=0.2).

A problem that arises in this case is that only one of the participating d.o.m. values dominates, by assigning its value to the whole decision criterion. This way eliminates the contribution of the other d.o.m. values. In the literature several other functions have been proposed for handling the logical operations on observations that have partial membership on more than one set. These functions are more complex to compute and interpret, however they provide more expressive and accurate results. One of those is the *energy metric* (Gupta et al. 1988), in which k fuzzy sets (A_1, A_2, …, A_k) defined over some domain C, create a new fuzzy set E with membership function given by:

$$MF^F{}_E(z) = \sum_{i=1}^{k} \left[MF^F{}_{Ai}(z) \right]^q \tag{4}$$

where q is a positive integer. By applying this equation (e.g., for q = 2; quadratic measure) the big weight values (d.o.m.) are amplified, while the small values are nearly eliminated. Assuming the previous example, the overall measure characterizing each individual location (l) of a region, regarding level ground *and* dry land using the energy function, is given by:

$$MF^F{}_{level-dry}(l) = \left[MF^F{}_{level}(l) \right]^2 + \left[MF^F{}_{dry}(l) \right]^2 \tag{5}$$

Notice that the energy measure derived by the previous formula should be normalized in the fuzzy domain [0,1].

Eq. 4 is more flexible and expressive than Eq. 2, because its overall measure is explicitly affected by the d.o.m. values of individual locations on both fuzzy sets A and B. Specifically, assume two individual locations L_1 and L_2 with the following d.o.m. values: L_1 {d.o.m. for level = 0.8 and d.o.m. for dry = 0.4} and L_2 {d.o.m. for level = 0.6 and d.o.m. for dry = 0.4}. The overall measure provided by Eq. 2 is 0.4 for both locations, while that derived by Eq. 5 is 0.80 for L_1 and 0.56 for L_2. Clearly the energy metric provides an *ordering* of the two locations. It says that L_1 satisfies better the two criteria posed by decision-makers (i.e., level and dry land). This feature of energy metric is very beneficial for decision criteria which combine multiple sets and lexical values, while ordering of the qualified entities (i.e., individual locations) is required (e.g., find the five most level and dry locations of a region).

4 The Data Model

An ideal GIS should provide a framework to represent all six dimensions of geographic entities (Section 2) in a compact and consistent manner. Commercial systems focus on spatial and thematic dimensions of entities and ignore or handle superficially the rest of dimensions. This is due to the inefficient data model they implement.

Following, a data model able to accommodate all six dimensions of geographic entities is presented. This model extends a spatial model, first introduced by Tomlin (1990). Tomlin's model treats in an elegant manner both single and related collections of spatial entities; and can be used to define spatial operations independently on the fundamental models available in literature (Aronoff 1989, Worboys 1995): the *vector* model and the *raster* model. In a previous study (Stefanakis and Sellis 2000), this model has been extended to accommodate the temporal dimension of geographic entities, while in a separate study (Stefanakis et al. 1999, Stefanakis and Sellis 1999), it is shown how this model and the associated operations can be extended to handle the uncertainty of geographic entities by applying fuzzy set methodologies.

The proposed model is defined as follows. Geographic entities are represented as a set of locations in space-time with a set of properties characterizing those locations. Depending on the application, space can be handled as two- or three-dimensional domain. For simplicity, this study examines the 2D geographic space. Since time can be considered linear, space-time is represented in a 3D modeling space. Two dimensions are reserved to describe the geographic location of entities and one to describe their temporal status.

The thematic data associated to geographic entities and characterize them depend on the application domain. *Fuzzy set methodologies* (Section 3) are recruited to provide a more realistic representation of thematic dimension. Specifically, the application domain defines a set of *k* lexical values characterizing a phenomenon *P*. Each individual entity *E*, related to that phenomenon, is attached *k* degrees of membership (d.o.m.), one for each lexical value. Those degrees are expressed in the range [0,1], where 0 indicates no-membership and 1 indicates full membership.

The data model can be viewed as a hierarchy of data (Figure 2). At the highest level, there is a space-time cube, which is called *map*. The map is a library of space-time cubes, called *thematic layers*. Each thematic layer refers to a theme characterizing geographic entities. Each thematic layer consists of a set of space-time cubes, called *lexical layers*, all of which are in registration (i.e., they have a common space-time coordinate system, which coincides with map system). Each lexical layer accommodates the d.o.m. value of each *individual space-time location* in the lexical value characterizing a theme. A space-time location is the smallest spatio-temporal unit of interest (and defined the generalization level) for the corresponding layer in the application domain. It is defined by the granularity in both space and time and can be generalized of specilized by applying appropriate operations. All individual locations with: a) a d.o.m. value within a specified range [a,b], b) geographic location within geographic space S, and c) temporal location within temporal space T, constitute a *space-time zone*.

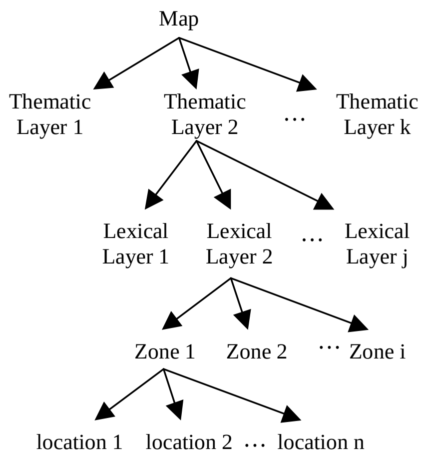
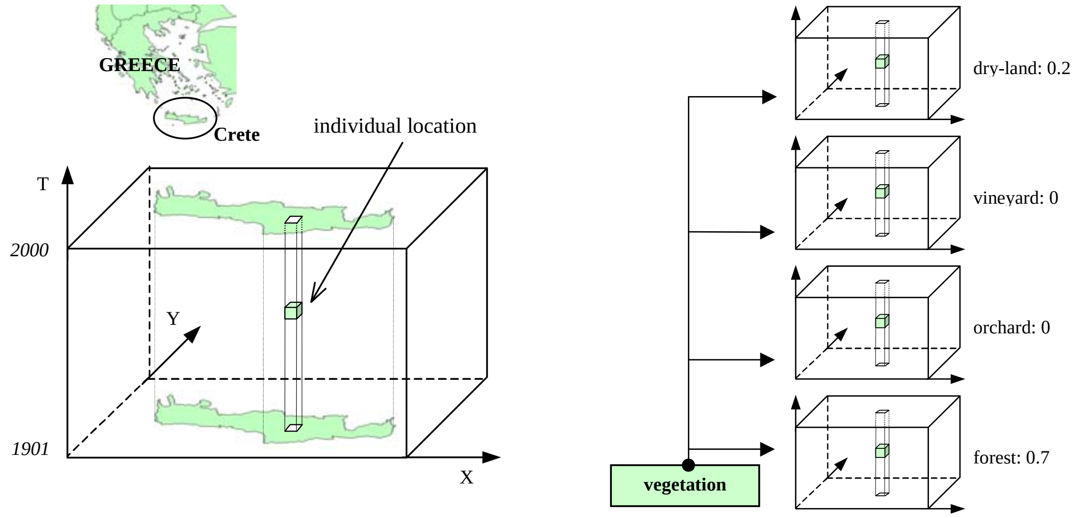


Fig. 2 The hierarchy of data.

Geographic entities of interest are represented by space-time zones and identified uniquely. Therefore, each entity is represented by a set of individual locations. For instance, a forest is represented by a set of spatio-temporal locations, consisting the corresponding spatio-temporal area (possibly with holes); a road network is represented by the set of spatio-temporal locations forming the roads of an area; a house is represented by one or more spatio-temporal locations (depending on the generalization level). It is important to notice that a geographic entity may change form or attributes over time and the proposed model may represent this information. On the other hand, the limitation of representing both spatial and temporal dimensions of continuous geographic phenomena with sharp boundaries (Burrough and Frank 1996).

An example of a map might be Crete Island in 20th century (Figure 3a). Example thematic layers are vegetation layer, land-use layer, road network layer, hydrology layer, hypsography layer, etc. Vegetation layer consists of the following lexical sub-layers (Figure 3b): dry-land, vineyard, orchard-trees, forest, etc. An individual location (X,Y,T) in lexical layer forest is assigned a d.o.m. value for forest, e.g., 0.7, which means that the location (X,Y) is forest with a degree of belief 0.7 at time T. All individual locations in forest cube with a d.o.m. value in the interval [0.7,1] and T = 1960, compose a zone (composite entity), which represents the forest areas in Crete Island in year 1960 (with a degree of belief greater than 0.7).



(a) (b)

Fig. 3 Example of map (a). The vegetation layer and its lexical layers with the d.o.m. values assigned to an individual location (b).

5 Generation of the Map

The generation of a new map is a composite procedure, which consists of the following steps (Figure 5). First, the application needs are examined and the scope of the map is well recognized. Then the parameters of the map are determined. That is, the set of thematic layers is defined, along with the lexical values per layer and the granularity in both space and time dimensions.

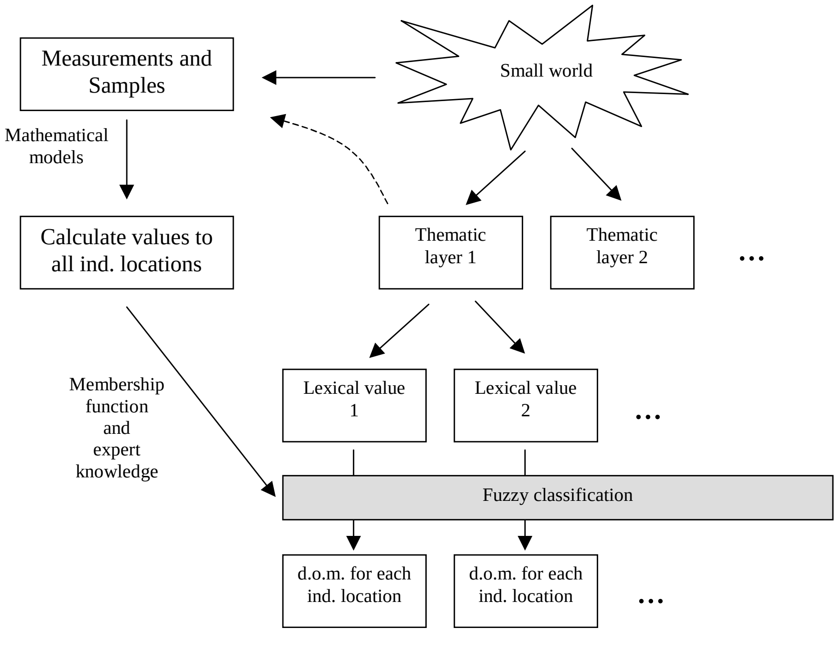


Fig. 4 Generation of a map.

Following the data collection is performed. Data collection includes discrete measurements of continuous phenomena (samples) in both space and time. Mathematical models are then applied to calculate of values for each phenomenon in all individual locations (based on measurements) per layer. Finally, d.o.m. values are computed, by applying appropriate functions and expert knowledge, for the predefined lexical values.

6 The Operation Model

There is no standard algebra defined on geographic data. The set of operations available in GIS varies from one system to another and heavily depends on the application domain. However, their fundamental capabilities can be expressed in terms of four types of operations: a) programming, b) data preparation, c) data presentation, and d) data interpretation operations.

Data interpretation operations are those that transform data into information and as such they comprise the heart of any system for handling geographic data. Data interpretation operations available in GIS characterize (Tomlin 1990, Stefanakis and Sellis 1998): a) individual locations, b) locations within neighborhoods, and c) locations within zones and constitute respectively the three classes of operations: *local*, *focal* and *zonal* operations. Considering the data model introduced in Section 4, these classes can be defined as follows:

– *Local operations*: they include those that compute a new value for each individual location on a layer as a function of existing data explicitly associated with that location (Figure 5b).
– *Focal operations*: they compute new values for each individual location as a function of its neighborhood. A neighborhood is defined as any set of one or more locations that bear a specified spatio-temporal distance and/or topological or directional relationship to a particular location (or set of locations in general), the neighborhood focus (Figure 5c).
– *Zonal operations*: they include those that compute a new value for each individual location as a function of existing values associated with a zone containing that location (Figure 5d).

All data interpretation is done in a layer-by-layer basis (Figure 5a). That is, each operation accepts one or more existing layers (cubic lexical layers) as input (the operands) and generates a new layer (cubic lexical layer) as output (the product), which can be used as an operand into subsequent operations.

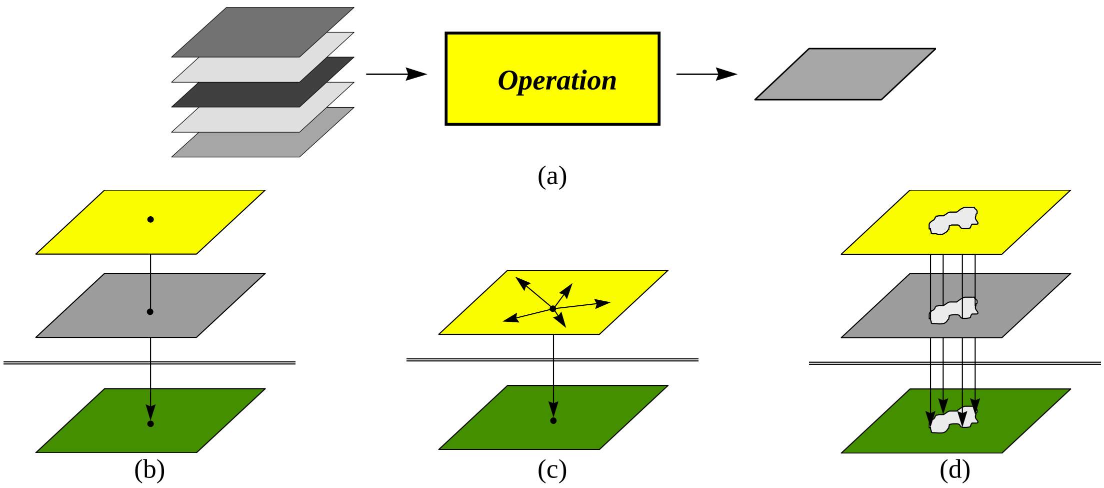


Fig. 5 Data interpretation operations (imagine level layers as space-time lexical cubes)

In traditional spatial information systems, such as GIS (Stefanakis and Sellis 1998): local class includes classification, recoding, generalization, overlay (spatial join) operations; focal class includes window, point queries, topological queries, interpolation, buffering, surfacial, connectivity operations; zonal class includes spatial selection, measurement operations.

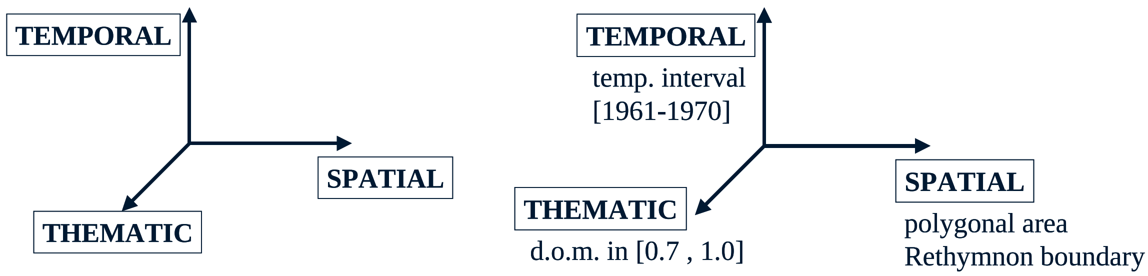
In a system adopting the proposed model and handling all six dimensions of geographic data, the operations above are extended in space-time domain while thematic data incorporate the measure of uncertainty.

The scope of this paper is not to provide an extended description of operation model. In contrary, the discussion is limited in a couple of basic operations in order to show generally the reasoning functionality of the proposed framework. In future studies specific applications will be examined and the required operations will be described in detail.

4.1 Select operation

This operation highlights all individual locations in a cubic layer, which satisfy a set of criteria. This set has three dimensions: a) spatial, b) temporal and c) thematic (Figure 6a). The spatial criterion refers to the spatial location of individual locations. The temporal criterion refers to the temporal location of individual locations. Finally, the thematic criterion refers to the d.o.m. values characterizing the individual locations in the cubic lexical layer.

An example select query is “find all forest regions in municipality of Rethymnon during the decade 1961-1970”. The corresponding operation might get as input the forest lexical layer of Crete and provide as output, all individual locations which (Figure 6b): a) spatially overlap Rethymnon region, b) temporally overlap the period of time 1961-1970, and c) have a d.o.m. value for forest grater than 0.7 (criterion posed by experts on the application domain).



(a) (b)

Fig. 6 Select operation example.

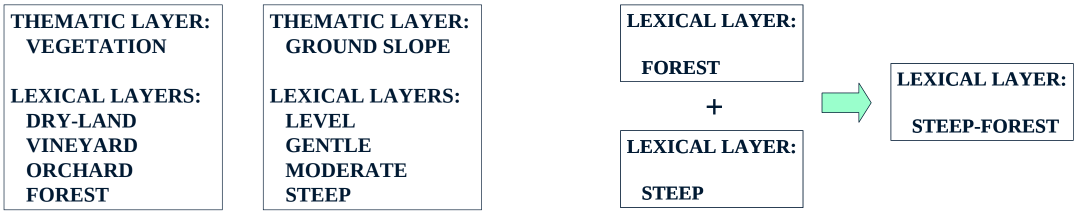
Obviously, in this example, if the temporal condition is degenerated into a point in time (e.g., “today”, or 1/1/1961) all highlighted individual locations (forest regions) lie on a plane which is parallel to the (X,Y) plane and correspond to a temporal snapshot.

On the other hand, in the same example, if the thematic condition is removed and the spatial condition is focused on a point in space, the result consists of a set of individual locations, which lie on a column in the cubic layer (same X,Y) and provide the changes of d.o.m. values assigned to that point in space, as regards to forest, over time.

4.2 Overlay operation

The overlay operation is commonly applied in GIS. It is analogous to join operation in conventional database systems, and is defined as the generation of a new layer and the assignment of attribute values to its individual locations resulting from the combination of the corresponding values in two or more existing layers.

An example overlay operation might get as input the cubic lexical layer forest (from vegetation layer) and the cubic lexical layer steep-land (from the ground-slope layer) and provide as output the cubic steep-forest layer (Figure 7). The later is a cubic layer whose each individual location has a d.o.m. value characterizing both how steep and forest it is. Performing a fuzzy operation and normalizing the result in the fuzzy domain achieve this.



(a) (b)

Fig. 7 Overlay operation example.

5 Conclusion – Future Research

This paper introduces a framework with enhanced capabilities in both representation and reasoning of geographic data. All six dimensions of geographic entities are incorporated in the data model and handled in the operation model. Several issues remain open and constitute the subject of future research. Specifically, the following topics are under study:

- Definition and implementation of a rich and compact set of operations for geographic reasoning.
- Extensive theoretical and experimental study on the choice of appropriate parameters for specific application domains. Those parameters include the granularity of space and time, the membership functions to simulate geographic phenomena, etc.
- The design and implementation of a prototype system for a specific applications domain.

References

Aronoff, S. (1989). Geographic Information Systems: A Management Perspective (WDL Publications).

Burrough, P.A., and Frank, A.U., 1996, (eds) *Geographic Objects with Indeterminate Boundaries* (Taylor & Francis).

Fischer, M.M. (1994). Expert systems and artificial neural networks for spatial analysis and modelling. *Geographical Systems Journal*, Vol.1, pp. 221-235.

Gupta, M.M., Knopf, G.K., and Nikiforuk, P.N. (1988). Sinusoidal-based cognitive mapping functions. In [Gupta and Yamakawa, 1988], 69-92.

Gupta, M.M., and Yamakawa, T., (1988). (eds), *Fuzzy Logic and Knowledge Based Systems, Decision and Control* (North-Holland).

Leung, Y., and Leung, K.S. (1993). An intelligent expert system shell for knowledge-based GIS: 1. The tools, 2. Some applications. *International Journal of Geographical Information Systems*, Vol.7, pp. 189-213.

Peuquet, D.J. (1984). A conceptual framework and comparison of spatial data models. *Cartographica*, 21(4), pp. 66-113.

Stefanakis, E. (1997). Development of intelligent geographic information systems. *Ph.D. Dissertation*. National Technical University of Athens, Greece.

Stefanakis, E., Vazirgiannis, M., and Sellis, T. (1999). Incorporating fuzzy set methodologies in a DBMS repository for the application domain of GIS. *International Journal of Geographical Information Science*, Vol.13, (7), pp.657-675.

Stefanakis, E., and Sellis, T. (2000). Towards the design of a DBMS repository for temporal GIS. In ESF-GISDATA Series Book: *Life & Motion of Socio-economic Units* (eds A.U. Frank, J. Raper and J.-P. Cheylan) (Taylor & Francis).

Stefanakis, E., and Sellis, T. (1998). Enhancing operations with spatial access methods in a database management system for GIS. *Cartography and Geographic Information Systems*, Vol.25 (1), pp.16-32.

Stefanakis, E., and Sellis, T. (1999). Enhancing a database management system for GIS with fuzzy set methodologies. In *Proceedings of the 19th International Cartographic Conference*, Ottawa, Canada.

Tomlin, C.D. (1990). *Geographic Information Systems and Cartographic Modeling* (Prentice Hall).

Worboys, M.F. (1995). *GIS: A Computing Perspective* (Taylor & Francis).

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8(4), 338-353.